IASSIST QUARTERLY

Volume 9 Number 3/4 Fall/Winter 1985

FEATURES

3 **Public Data In Use: A Case Study of Ireland**
By John Blackwell

14 **Prospects and Problems In The Use of Government Generated Data in India**
By N.K. Nijihawan

20 **Building Regional Data Files in the Federal Republic of Germany**
By Edwin Ferger

25 **A Network System For the Flow of African Administrative Information Materials**
By Edward Seth Asiedu

34 **Access To Unpublished Social Science Data in Government Departments in the United Kingdom**
By D.A. Clarke

DEPARTMENTS

37 New Products
41 The BBC Domesday Project
42 Comparative Projects
45 World Election Surveys

Printing compliments of The Rand Corporation

IASSIST QUARTERLY

Volume 9 Number 3/4 Fall/Winter 1985

FEATURES

3 Public Data In Use: A Case Study of Ireland
By John Blackwell

14 Prospects and Problems In The Use of Government Generated Data in India
By N.K. Nijihawan

20 Building Regional Data Files in the Federal Republic of Germany
By Edwin Ferger

25 A Network System For the Flow of African Administrative Information Materials
By Edward Seth Asiedu

34 Access To Unpublished Social Science Data in Government Departments in the United Kingdom
By D.A. Clarke

DEPARTMENTS

37 New Products

41 The BBC Domesday Project

42 Comparative Projects

45 World Election Surveys

Editorial Information

The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers.* Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. Bibliographic references for numeric social science data files: suggested guidelines. *Journal of the American Society for Information Science 30(2):77-82, March 1979.* If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor:

Walter Piovesan
Research Data Library
W.A.C. Bennett Library
Simon Fraser University
Burnaby, B.C., V5A 1S6 CANADA
(01)604/291-4349 E-Mail: Piovesan@SFU.MAILNET

Book reviews should be submitted in duplicate to the Book Review Editor:

Kathleen M. Heim
School of Library and Information Science
Louisiana State University
Coates Hall, Room 267
Baton Rouge, Louisiana 70803 USA
(01)504/388-3158

Key Title: Newsletter - International Association for Social
Science Information -Service and Technology
ISSN - United States: 0739-1137 Copyright © 1985 by IASSIST. All rights reserved.

Public Data In Use: A Case Study of Ireland

by John Blackwell[1]
Resource and Environmental Policy centre,
University College, Dublin

Introduction

This paper is concerned with the policy issues which arise in enabling the public data which are most useful for research purposes to be disseminated in a timely fashion. Case study material from Ireland is drawn on, outlining problems which have arisen in identifying and in meeting the needs of users, and the extent to which it is possible to meet the requirements at a time when there is no prospect of an increase in real Government expenditure on the provision of statistics. Particular examples in the areas of earnings, employment and social welfare statistics will be given. The extent to which administrative records can be a fruitful source of data, and can substitute for purpose-built surveys and censuses, especially in view of the cost constraints, is assessed. At the end of the paper, some policy issues are posed.

The background – data provision in Ireland

In Ireland there is a centralized Central Statistics Office (CSO), whose output ranges from population and vital statistics to labour force, employment and unemployment statistics, industrial production, prices, earnings and hours worked in industry, trade, transport and distribution. Many of these statistics are produced by means of censuses and surveys. The methods of dissemination used by the CSO vary from published volumes to regular mimeographed series (issued usually monthly or quarterly); data on Ireland which appear only in publications of international organizations such as Eurostat and OECD; small area data such as from the Census of Population and from the Census of Agriculture which are disseminated directly to the users; and special analyses which the CSO does in answer to specific requests e.g., from the labour Force Survey and the Household Budget Survey, which are usually not charged for.

However, in recent years there has been an increase in the amount of data available from public sector bodies outside the CSO, which in the main complement the existing range of data from the CSO. The monetary and public expenditure data, produced by the Central Bank and the Department of Finance respectively, can be put with the national accounts as the basic raw material for macroeconomic analysis. Much of the information on education and manpower flows and science and technology, on activities such as transport, energy, and tourism, on education, health, housing: and social services, and on taxation are produced by a variety of bodies in the public sector outside the CSO. In a number of cases, these data are produced by government departments (e.g. Education and Social Welfare); in other cases the data comes from government agencies or state bodies (e.g., in the case of science and technology statistics).

---

[1] Presented at IASSIST/IFDO International Conference May 1985, Amsterdam

---

Problems of resource allocation

There has been no increase in the volume of resources provided to CSO in recent years. rather, a decline. While there is no coherent account of the amounts of resources which are put into provision of statistics outside CSO, these are unlikely to have increased. The budgetary problems in Ireland rule out any prospects of an increase in resources for statistics. Indeed, one of the elements which worked to the advantage of data users in Ireland in recent years was Ireland's entry into the European Economic Community. This meant that EEC regulations and directives on statistics became binding. This provided, for example, a much needed set of statistics on the distribution of earnings through the 1979 Structure of Earnings Survey. As a reflection of EEC budgetary pressures, a proposed 1986 survey will not now take place, and there is no domestic survey which can fill the gap.

Especially at a time of pressure on the government budget some means of ranking the statistical output by reference to the intensity of demand, is required. This is very difficult to achieve, for the following reasons:

- There is a heterogeneous body of users, ranging from government to research institutes, individual researchers and private sector firms. The implied value of attributes such as timeliness, frequency, accuracy of estimate, level of disaggregation, will vary among these groups of users. Many private firms, and those in the financial sector, put a premium on timeliness. Hence, as a result of timeliness problems in the case of the Irish Census of Industrial Productions (problems which are currently being righted, slowly), there has been minimal use of industrial statistics by business firms. There are other reasons for this neglect of industrial statistics on the part of the private sector. Firms will typically want highly disaggregated data in order to answer questions such as market share or productivity change for a particular product. In a country such as Ireland, where one or two firms may dominate some sub-sectors, there often there can be no possibility that CSO would be able to provide the data for reasons of confidentiality. Hence, the popularity among business users of the foreign trade statistics, which are quite disaggregated and are more timely than production data.

- Related to the preceding point, it is difficult to establish the nature of the "trade-offs" which exist between improvements along these different dimensions of statistics. For example, timeliness could be improved by sacrificing some accuracy of estimate. For some firms, such as those in the financial sector, timeliness is crucial. For them, it is often more important to have an early estimate which is subject to later revision, than to have a purer number which arrives after a long interval. By contrast, in the case of social policy, the time-scale is different. Here, the analysis of large data sets, gathered at longer intervals, concerned more with details at household or family level, is often needed. With regard to frequency, there is a danger that in a perverse way, an over–concentration on frequency could slow up decision–making, especially if estimates are subsequently revised to a marked extent. In the case of national accounts data – where there is already a time–lag due in part to slowness in provision of local authorities' data – there has in Ireland been a good deal of data revision long after the event.

- The price mechanism is not used to ration demand so that those uses of highest value (or which use up least scarce resources) are effected. That is partly because information has many of the attributes of a "pure public good". This is, information which is made available to one person (through, for example published sources, in the widest sense) can be made available to the community as a whole, and one person's consumption of information does not "take away" from the consumption of another person. (Of course, this must be qualified: information can be withheld from certain consumers and can be bought and sold on the market.) Moreover, a number of the statistics are provided at close to zero prices, which gives rise to difficulties in assessing the intensity of preferences among actual and potential users.

In summary, therefore, a key issue is how many more, or how many fewer, resources should be devoted, at the margin, to producing particular data series. This is, in a sense, a problem of providing a surrogate for the market system in guiding resource allocation to the provision of public statistics. Difficulties abound in assessing the intensity of user preferences, partly because of the various preferences of the different users. Furthermore, it is difficult to get people to reveal their preferences, a reflection of the 'public good' aspect of statistics. Users will be tempted to overstate their preferences, in the knowledge that if information is provided by somebody else, they in turn can benefit. In other words, people can obtain information at no cost to themselves, even though, if they were confronted with the stark alternative of "pay up or do without the information" they would be prepared to pay something.

The "classic" perception of the way in which information is gathered and organized has often centred on the following sequence by which decisions are made: identification of a problem; collection of data to illuminate that problem; identification of feasible solutions; choice of the best solution from the alternatives. However, in practice, this logical sequence is not always followed. Information may be gathered only after a proposal has been decided on, either to monitor the implementation of policy, or to co–ordinate activities across Departments, or to serve as a back–up for particular cases which a public body may wish to make.

Why does the classic model not apply? In part it reflects the fluctuating pressures for decisions, often made under time pressure. There is inevitably a time lag before information can be gathered, by which time decisions may have already been made. In part it reflects the lack of a clear set of signals from the users of data to the producers. In Ireland, the CSO has found it difficult to get users to articulate their requirements an unambiguous way. While there has been a generalized feeling of dissatisfaction among users, there has been no formalized method whereby the reactions of users can be channelled to the producers. At mid–1984, for the first time, a one–day statistical users' seminar was held in order to help communications between users and producers. Even here, there was a disappointing turnout from private sector users. CSO does consult with major users, but this consultation gives little weight to non-institutional users.

It is almost trite to observe that the demand for statistics is a derived demand. Yet it should be noted that these derived demands are governed by the ecomic structure and by the way in which policy analysis and research occurs. Thus, for example, most social science research in Ireland occurs in large research institutes. In the absence of a funding body for social science research, and given a weak tradition in team research in the social sciences, the countervailing power of the lone researcher, not to speak of ability to use large data sets, may be weak. Indeed, for some uses such as macro-economic models, the data supply/demand position is more or less one of one monopoly producer (CSO) and one or two large users such as the Economic and Social Research Institute.

Another example is the use, in other countries with more decentralized of government, of sub-national statistics to govern the spatial allocation of public resources: with the degree of centralization in Ireland, this use is virtually absent. A final example: the more rapid the pace of social change (such as women's participation in the labour force), the greater the likelihood that the full censuses will (again, given current methods) lag behind what would be needed to monitor change. This has implications for the potential use of sample surveys or else of simpler censuses.

There will tend to be a chronic excess of user demands over the supply of statistics (given existing resources and methods). The question arises whether it would be possible to decide on allocating resources to statistics by using formal criteria such as cost-benefit analysis. The answer is "no". While it is possible to estimate costs, the diffuse nature of the uses of statistics would make the estimation of benefits extremely difficult. Except in a few particular cases, the monetary value of statistics will be almost impossible to compute. Moreover, given the links between information, policy and "background" knowledge of the world, the potential returns from statistics may be greater than the actual returns if there is potential to improve the linkage between public policy-making and the provision of statistics. This can be illustrated by reference to the Irish Census of Population. The cancellation of the 1976 Census had some direct effects on other data provision – it led to increasingly inaccurate estimates of population and of the labour force and consequently of employment and of national output. At the same time, there seems to be almost no accounting for the individual decisions, including those in physical planning, which have been affected. But the measure of these effects would be different with a different type of public administration, one that was more responsive to changes in economic and social information. In this sense there are interaction effects which complicate any attempt to ascertain a pure return to statistics.

It is difficult, in fact, to point to particular policies in Ireland which were critically influenced by the provision or absence of certain data. There are a number of reasons for this. First, the provision of data *per se* is best seen as contributing to the understanding and explanation of economic change and social processes. Second, even if a starkly empirical view were (mistakenly) taken – that the matter begins and ends with gathering "the facts" – the dissemination of data would work its effects in a diffused way, with time lags. Third, there are usually time pressures on policy-makers – research findings and associated data must be speedily available, and this is not always possible. Fourth, it is difficult to predict data needs ahead of time – the focus of policy attention tends to change over time, while the provision of statistics lags behind. Finally, there are insufficient contacts between researchers and policy-makers (including lack of effective dissemination of findings) and there is certainly too little interaction between the three groups of policy-makers, researchers and producers of data.

However, it is possible to get a feel for the value of statistics by asking in any particular case:

- what are these statistics for?
- who is using them?

In anwering these questions, due account should be taken of the nature of the benefits, many of which are indirect, rather than relating to specific uses by those who decide on public policy. And a number of benefits relate to users outside Government.

In particular, there is need to concentrate on decisions at the margin - what additional information should be provided, and what should be dispensed with? This would enable the identification of those data which are most crucial to policy decisions. It would also enable the recognition of those pieces of information which are redundant, or where a lowering in frequency or a degeneration in timeliness could be tolerated, especially if this meant that resources would be freed to devote to the provision of information for which there would be a high return. In Ireland it has, however, been extremely difficult to identify "redundant" data series. This may in part reflect a system which is producing the "minimum critical mass" of statistics. It may also reflect a "ratchet effect" whereby, once a series is put in place, any attempts to discard it are met with opposition.

Scope for Co-ordination and Rationalizaton

While the collection of statistics by a wide variety of public bodies outside the CSO has filled a number of gaps in knowledge, it has given rise to a number of problems.

- No explicit co-ordination of this output of statistics occurs, nor is there any indication of the amount of resources which go into this activity. There is no "statistics budget" for the public sector as a whole, with an associated set of outputs. CSO itself has no "leverage" over the gathering of statistics by other public sector bodies. The activities of CSO could be greatly influenced by the data gathering of other bodies, either because they complemented the work of the CSO, or because the CSO could use the data for purposes such as national accounts compilation, or because data could be used either as a means of verification, a sampling frame, or a register for coding purposes.
- There is no quality control with regard to non-CSO statistical output. This has in part reflected the way in which data collection began - sometimes as a by-product of filling internal planning needs and the need for co-ordination, sometimes as a result of implementing administrative schemes. Unpublished data are available from an employment survey conducted annual by the Industrial Development Authority, from a quarterly consumer survey undertaken by the Agricultural Institute for the EEC, and in an information system on public sector employment in the Department of the Public Service. The first and last cases exemplify data sets which have been compiled primarily for internal

management purposes.

- There are no principles of access to non-CSO data. While there has been some dissemination of the Industrial Development Authority and Department of the Public Service data, this has been as a by-product of secondary analysis by "outside" workers. The Industrial Development Authority data have been used as a source for information which, in the absence of CSO industrial employment data on a "component of change" (job gains, job losses) basis, cannot be obtained from other sources. They could, in part, be substituted by CSO data if the basic CSO data were analysed in a form which facilitated the "component of change" work. Government departments hold a great deal of information which is of interest to policy analysts, some of it in the form of raw data. A certain furtiveness in even disclosing the presence of data seems to be endemic. The ultimate barrier is when a Department or agency does not release a body of data. For example, twice since 1973, the results of housing conditions surveys, undertaken on behalf of the Department of the Environment have been kept within the Department. This is a notable case, because the information would complement the census data on housing and, in the areas of unfitness and obsolescence, provide information for policy which is more relevant than could ever be obtained from the census.

- While there are many *ad hoc* surveys conducted by Departments and other bodies, surveys conducted by the Economic and Social Research Institute for the Institute work, and surveys conducted by other research institutes and by individual researchers, none of this body of material is pulled together in one place. Nor is it even catalogued. There is no central user's guide to the body of data.

- There is danger that the current round of computerization which is occurring in public sector bodies will result in further fragmentation. There is, for example, a danger that computerization among local authorities will miss out on the possibility of providing the means of a data bank. To take another example: the Department of the Public Service is to expand its personnel information system. Ideally, in this process, the CSO should be consulted and compatible codes for occupations and other entities should be used.

This leads to the issue of whether there would be potential for rationalizing the data-gathering on Irish by State-sponsored bodies. A number of these bodies - national training agency, the Irish Export Board, the Industrial Development Authority, the Institute for Industrial Research and Standards, and the National Board for Science and Technology - compile data on Irish firms, as does the CSO itself. On the surface, rationalization could contribute to the problems of non-response and of slowness in response, which in part account for the time lags in the issuing of CSO industrial statistics. This may be the case up to a point, but it has to be qualified in view of the client relationship which exists between many of these firms and the State body in question, which is likely to facilitate response. A more compelling reason for seeking some rationalization is the richer potential for cross-tabulation which would ensue, as at the moment this is limited to the variables on each individual inquiry. Moreover, there would arise a common set of standards with regard to methods of compilation and of analysis, which would ideally be in accord with CSO categories as regards coding. There would, of course, remain a sub-set of data which would be confidential to individual agencies.

Some co-ordinating role is needed in relation to the output of statistics from sources other than the Central Statistics Office. There are a number of ways in which the desired co-ordination of non-CSO sources could occur. One would be through expanding the responsibilities of CSO to include the co-ordination of all public statistics. Another would be a co-ordinator of non-CSO. The co-ordinator could be in a Government department, or outside the departmental system, or could work through an inter-departmental committee.

The main objectives of the co-ordinating function would be as follows:

- To encourage public sector bodies to consider the statistical possibilities of administrative records and to consult on this with bodies such as the CSO. This is especially important at a time when new computer facilities are being installed.
- To ensure that statistics which are of importance to policy-makers but which are of little or no importance to the collecting agency, are produced. This may at times require that such a subordinate objective be given explicitly to the agency.
- To ensure quality control, i.e., some minimum standards and compatibility between individual data sets. The more standard classifications and definitions are used, the lesser the likely burden on respondents as their records would be in conformity with all requirements.
- To examine the potential for rationalization in the data-gathering activities of State bodies. This could lead to richer possibilities of data analysis, as it might be possible across the headings which are contained in the existing surveys. There are two ways by which this could occur: through an "umbrella" survey, or through the linking of records. The latter would be much more difficult to achieve. Some move towards an "umbrella" survey would be the most fruitful route.
- To consider the use of these statistics as a means of verfication of sampling frames or registers for coding purposes.
- To take part in a revised and explicit system of resource allocation for statistics, influencing the allocation of resources and pointing up the areas where attention needs to be focused.
- To fulfill some gatekeeping functions with regard to the increasing volume of surveys of Irish firms by those in the public sector.

Among the cases which are candidates for action are manpower and energy. Another instance is the road vehicle file of the Department of the Environment which is computerized but which is not being used to produce data on the vehicle fleet by age, make, type of fuel used. Perhaps the prime candidates are the Department of the Public Service data on public sector employment (mentioned above) and Revenue Commissioner's data derived as a by-product of tax collection. This leads to the next section, on the potential for using administrative records.

Using Administrative Records

In part because of the budgetary problems which face us, the use of administrative records as sources of statistics is an attractive option. Limited resources can be devoted to the gathering of statistics by surveys; statistics based on administrative records can be collected at relatively low cost, in contrast with surveys. In addition, there has been increasing resistance on the part of respondents to provide information to the statistical office. (Indeed, this was the main reason the pre–1968 survey on the distribution of industrial earnings by the CSO was dropped). In addition, income data derived from administrative records have potential for being more accurate than income data derived from surveys, which are subject both to sampling error and to understatement of income. This potential is particularly evident in the case of the service sector where relatively little data are available on earnings and employment – even for the public sector, where the only readily available information is on civil service employment.

The potential for making greater use of administrative records as a data source, or as a basis for sampling, will increase as existing records become more computerized, thereby enhancing the potential for cross–tabulations. A number of departments and agencies are now apprised of the potential for decentralized collection and dissemination of data via micro–computers. The idealized picture, thus, is one where the contraints on resources which currently bind us are loosened through a combination of the use of administrative records allied with computer collection and analysis. There are a number of potential snags which could disturb this pleasant sequence of events.

First, the definitions and categories used in administrative records reflect the administrative process (e.g., entitlement of benefits, being in the tax net). They may not be ideal for statistical analysis, and many change over time. Ideally, statisticians would be involved at an early stage, either when new systems are being put in place or when manual records are computerized. One feels apprehensive that the current wave of computerization could lead to a rash of incompatible variables and codes. Experience to date with some of non–CSO sources–with their lack of compatibility with CSO data and changes over time in categories which impede time series analysis – does not inspire confidence that this can be avoided without some policy initiatives. The Department of Social Welfare data, which have a good deal of potential for providing data on social welfare recipients, are tied to benefit recipiency. Hence, it is difficult to explore problems of lack of take–up of benefit. And in the case of unemployment data, in Ireland (as in other EEC countries) a substantial minority of the unemployed (from the labour Force Survey) receive neither unemployment benefit nor unemployment assistance.

Second, much of policy analysis requires linkages – either to CSO data or to departmental data. An example would be the relationship between social welfare payments and patterns of work. Such linkages are either quite difficult to effect or can raise problems of confidentiality.

Third, the only aspects of administrative records which are reliable are those which are regularly used by the authority in question and are, therefore, kept accurate and up to date. Considerable clerical resources may be required to maintain administrative records for statistical analysis purposes (i.e., keeping designatory details up to date, elimination of duplication, etc.) and authorities have other priorities in mind. If administrative records are to be fruitful, they need constant attention. Important administrative details required for analytic purposes (e.g., designatory details such as business descriptions, occupation, age) are not usually complete nor up–to–date because they are not frequently used by the authority. This is the biggest handicap to the use of these records, and arises in particular in the case of the Irish Revenue Commissioners data which have the potential to provide earnings and employment data and to fill notable gaps in coverage, especially on the service sector.

Fourth, there is need for close coordination between statisticians and administrators in order that the needs of both administration and provision of statistics are met. Decisions have to be made about classifying data in cases where there may well be conflicts between the desires of different users. At times when schemes are designed, changed or computerized, there is need to be sensitive to the possibilities of using these records to provide statistical tabulations. This would mean consulting with statisticians in the design stage. If consultation were to occur too late, a costly redesign of systems could be required. In the past in the case of the Irish health sector, where there is a certain centralizaton of the functions of Health Boards, computer systems which were incompatible across Health Boards were acquired. While there has since been a rethink of information requirements within the Department of Health, this has involved too little contact with those with a planning function in government.

Fifth, a major constraint to using the Revenue Commissioners' data relates to objectives. This is linked both to the issue of resources and the need to use clerical resources. Understandably, the Revenue Commissioners regard their main purpose as getting in money from the tax system and would regard information provision as, at best, a subsidiary objective, or at worst having no justification at all. In the case of a tax with high collection costs per yield, such as the residential property tax, the minimum amount of information would be provided. From the point of view of the public weal, and from the point of view of cost-minimization for the State as a whole, there is need to use the Revenue Commissioners data. But there is no incentive for the Revenue Commissioners themselves to take account of this. Rather the opposite: if people are switched from revenue-raising in order to provide more information, the Revenue Commissioners will be in danger of receiving criticism for failure to maximize tax revenue.

If the basic data were "clean", it would take little resources to produce the required data, by using the automatic data processing facilities which Revenue has. The biggest obstacle to producing tables which would help policy-making outside the specific remit of Revenue is the need to deploy clerical workers to "clean" data with regard to designatory codes. This would especially be the case as Revenue is prepared to consider the provision of a subset of basic data on magnetic tape to outside agencies - with the individual identifying characteristics removed. Users could then analyse these data using either a software package or a custom-built program. This would be a particularly valuable facility.

There will of course, always remain areas of inquiry where there is no substitute for survey or census data. An example arises in the case of labour force statistics.

It is manifest that existing data are often inadequate to answer the analytical and policy questions which arise. This is partly because most of the labour force data are given in terms of stocks. Many of the questions raised need an analysis of flows between states. Another reason for difficulty is that existing data come, in part, from administrative records which are tied to benefit criteria, which differ over time, and in part from self-description of states which are inherently ambiguous, such as "unable to work due to disability". An instance of this from the Irish Labour Force Survey is that in 1977 a category "unable to work due to permanent sickness or disability" was introduced, but led some people who described themselves as "unemployed" in 1975 to opt for this new category. There is an interplay between disability, unemployment and disability, and between retirement and disability. If one is to capture the impact of social security payments on the labour market, there is need to complement the existing stock data with flow data.

Ideally, flow data (shown in Table 1) by sex would be required. In this table, the intersection of a row and a column gives the flow between states. From this table, a table showing transition rates (for, say, one or two years) between labour force states could be constructed. This would show entry rates into a particular state from all types of states, together with exit rates, and continuation rates (the proportions of people who remain in a particular state between one period and another).

In Ireland there is already one survey which in part can be used to provide flow data: an annual survey of school-leavers undertaken for the Department of Labour. This is an example of a survey which would provide the answer to a number of policy questions if the basic data were accessible.

Policy Issues

In summary some of the policy issues which arise are:

- There is need for some form of coordination (with no coercion?) of non-CSO data with CSO data, in order to ensure quality control, i.e., some minimum standards and compatibility between individual data sets. There is also scope for some gatekeeping functions with regard to the increasing volume of surveys, including surveys of Irish firms, by those in the public sector. And there could be some rationalization of the existing surveys of Irish firms, leading to richer possibilities for data analysis.

- Should some minimum standards be set in relation to data which are collected as a by-product of government contracts, including their deposit in a central location?

- What forms of surrogates for the market system to identify user demands ("needs"?) can be developed - e.g., some mechanism for feedback and complaints, for wide consultation and receipt of suggestions at an early stage of the design of inquiries? For example, could there be greater use of charges in order to get an indication of the implicit value which different users put on different types of statistics?

- What forms of users' guide are required for the existing mass of data which public bodies hold? And what ground rules on access to public sector data by outside users should be established?

- There is need for public sector bodies to consult with CSO on the statistical possibilities of using administrative records. Related to this, how can the danger that the increased diffusion of computers among public bodies, including the local authorities, will lead to the onset of incompatible data sets, which are in turn incompatible with CSO data, be avoided?

- Despite the above remarks on the potential for using administrative records, there are many issues of economic and social policy where there is no substitute for survey or census statistics at the level of the individual or of the household, especially where it comes to establishing relationships between variables. This is especially true in view of the increasing incidence of "non-standard" households such as lone-parent families and the many family policy issues which result. To what extent could existing surveys be used with additional questions being asked at marginal cost? In Ireland, there are many possible uses of a general household survey with a central base of questions and a varying sub-set of questions from year to year. This could provide information on education and on use of health services, for example. There is no possibility of adding to the Household Budget Survey (i.e., family - expenditure type), which is already overloaded.

- To what extent should a simpler Census of Population be conducted, with a shorter time lag for dissemination, leaving matters such as housing to sample surveys?

- To what extent should one try to force the development of intermediaries who would take primary industrial data and engage in "packaging" and marketing to industrial customers; are there economies of scale in this activity?

- Underlying the paper as a whole, there are the difficulties in deciding how much public resources to put into statistics. For example, non-CSO data will reflect producer preferences as to internal management needs or inter-departmental and inter-agency coordination: this may not be optimal from the point of view of society as a whole. ⌑

Schematic outline of type of labour market flow data which are required

Initial State	Final State (numbered as in rows)							
	1i)	1ii)	2	3a)	4	5	6	7
1. Employed								
a. full time								
b. part time								
2. Unemployed								
3. In partial retirement								
a. In the same job as when last fully employed								
b. In a different job								
4. Fully retired								
5. In labour force, sick								
6. Unable to work due to disability								
7. Not in labour force due to "home duties"								

Prospects and Problems In The Use of Government Generated Data in India

by N.K. Nijhawan[1]
Director, Data Archives,
Indian Council of Social Science Research (ICSSR),
New Delhi

This paper attempts to discuss the nature and extent of government effort, in India, in generating various kinds of socio-economic data and also to examine the problems faced by researchers in having access to such data. India is one of the few countries in the world which has such an elabourate, comprehensive and multilevel data-base. But it must be admitted that very few of these data are put to effective use. The government departments and its other organizations have a limited purpose and accordingly analyse the data only marginally. The researchers in the institutions and university departments who have interest, necessary skills and competence to use data in greater depth and dimensions do not have access to these data. Thus despite a 'data revolution' in the country, both policy makers and researchers are unable to make effective use of a valuable data base for various reasons. An attempt has been made to analyse some of the problems underlying this process and also to explore possibilities that exist for remedying the situation.

Historical background

The present statistical system in the country has evolved through centuries. It dates back at least to the sixteenth century, when the Mughal rulers had data collected and tabulated systematically for about four thousand geographical units of the Mughal Empire. A comparable system for that century may not be found in any other part of the world. The nature of tabulated data also indicates that the Mughals may have borrowed some of its important features from the system prevalent in an earlier period. It seems even the British rulers, in India, leaned heavily on the data system developed by the Mughals during the early period of their rule. The British, however, transformed, and in many ways enriched, the system as is evident from the fact that the first gazetteer was published in 1866, the statistical abstract of British India in 1868, and the first all-India population census conducted in 1881. In 1895, a statistical bureau was established at the Centre to coordinate agricultural and foreign trade statistics and to collect data on prices, wages and industry. A number of Committees and Commissions were set up from time to time to examine in detail the available statistical material with a view to improving its quality.

---

[1] Presented at IASSIST/IFDO International Conference May 1985, Amsterdam. The views expressed in this paper are those of the author and not those of the ICSSR.

Post Independence Developments

A real breakthrough in the official statistical system, however, came only after independence. A nucleic statistical unit called Central Statistical Organizaton (CSO) was set up in 1951. Subsequently, State Statistical Bureaux (SSB's) were set up in almost all the States. Following this statistical offices were set up, each in the charge of a district statistical officer. Like the CSO at the centre, the SSB's are entrusted with the task of coordination and publication of statistics collected by different government departments and district statistical offices.

At the instigation of, then, Prime Minister Jawaharlal Nehru, a National Sample Survey (NSS) was started in 1950 by the government to expeditiously collect data relating to all aspects of the national economy, on a continuing basis, using sampling methods. Beginning with the first survey (called a Round) in 1950–51, the National Sample Survey Organization (NSSO) is about to complete data collection for the 40th Round by July, 1985. Currently, NSSO has a large staff. Its field staff alone consists of about 5000 full time research investigators and field supervisors engaged in data collection.

The need for collection of relevant data for the purposes of National Five-Year Plans, as well as for meeting the demands of various international organizations, has been increasingly felt since independence. A large body of these data are collected by the government in addition to those that are usually collected for its routine and normal functions. It is appropriate to say that the country is witnessing data explosion since 1951. A brief mention of the nature and extent of socioeconomic data generated by the government during the past three decades or so is in order.

Nature of Data

Four decennial all India population censuses have been completed. Basic population characteristics for all the districts (approximately 365) are available in published form for all four censuses. Another extremely important tabulation of the Indian census which goes right up to the village and urban blocks called the primary census abstract (PCA) are available for each village/town in the district census handbooks. The census data also serve as the major source of basic economic information about the populaton. Besides data on population composition and distribution, data on fertility and mortality are collected under the vital registration system.

Considering the significance of agriculture in Indian life, information on agricultural holdings, their characteristics such as: size, form of tenancy, land utilizaton, production of crops, irrigation, livestock, prices, wages, etc., are being collected systematically by various government agencies. Data pertaining to land utilization, areas under various crops and area irrigated are almost continuously available in published form since 1884. Apart from the data collected from village records, NSSO has also collected data on land holdings, land use, agricultural production, cost of cultivation, etc., in its various Rounds. Besides, the NSSO carried out a special survey on land holdings, at the instigation of FAO, in its 16th and 17th Rounds to meet the requirements for the 1960 World Agricultural Census.

Unlike agricultural statistics, data on Indian industry is of relatively recent origin. Before independence, industrial statistics in the country, were extremely inadequate and of poor quality. After independence, legislation had to be passed to facilitate collection of statistical data relating to industries. A breakthrough came with the passing of the Collection of Statistics Act, 1953 and the Collection of Statistics (Central) Rules, 1959. An annual Survey of Industries (ASI) has been carried out since 1959, covering all the registered and licensed factories in the country. A number of government agencies periodically collect data on employment, output, earnings, etc., related to small scale and cottage industries. At present, the data relating to the industrial sector are available from at least twenty different sources of the government.

The National Sample Surveys are a rich treasure of data encompassing a large spectrum of socio-economic and cultural life of Indian society in the post-independence era. The NSS are multipurpose surveys covering several socio-economic and other subjects simultaneously. Comprehensive data have been collected, in various Rounds, on different aspects of agriculture including land utilization, cost of cultivation, rural indebtedness, rural labour etc. Likewise, a great deal of effort has gone into data collection in various surveys on consumer expenditure, demographic characteristics of sampled households, internal migration, employment and unemployment, housing conditions, trade, transport and rural electrification.

The quantity of data being collected through these surveys is quite voluminous. Just to give an idea, it may be mentioned that the sample size varies from Round to Round, but normally a survey covers about 5000 urban blocks including 200,000 sampled households. Sometime back, at the initiative of the Indian Council of Social Science Research, an estimate was made of the number of schedules/pages canvassed in various NSS Rounds. It was estimated that in the first nineteen Rounds alone (1950–51 to 1964–65) over two million schedules, comprising over nine million pages, had been canvassed. The large quantity of information collected during the course of all these years can thus be imagined.

Lack of perspective

The above account, which is purely illustrative in nature, clearly reveals the nature and extent of data being collected by the government. By any standards, India can claim to be one of the richest countries in the world in terms of data resources. But at the same time it cannot be denied that both policy makers and researchers are unable to make effective use of this resource, as a large part of the available data is put to a very limited use. The obvious question to ask is: why is that so? The single most important factor responsible for this sorry state of affairs is the government's lack of perspective in this regard. It may be worth noting that there is a disproportionate emphasis on data collection as opposed to data utilizaton. Emphasis continues to be on data collection rather than on its utilization and improvement of its quality. The main problem, therefore, is that the effort in data collection has multiplied manifold in the past 35 years, but the techniques of handling data continue to be quite rudimentary. Most of the data collected by the government is tabulated manually even today. This results in long delays, even in bringing out limited tabulated data. For instance, the results of the 31st NSS Round on rural electrification, conducted in 1976–77, were published as late as January, 1984. Likewise, although the summary results have been released, the detailed results of the 1978–79 Annual Survey of Industries are still being processed by the CSO. The last detailed results of ASI available for public use relate to the year 1973–74.

Nothwithstanding unusually long delays in the publication of government-produced data, scholars can make little use of them owing to high levels of aggregation in the published data. Most government agencies collect data at a micro level, such as, village, household or industrial unit, but these are not made available to researchers. The analysis of micro data at the village level would have been immensely useful to the researchers, but such information is not available to them. Only all-India and State-level results are normally published by the government.

Lack of Uniformity

Multiplicity of government agencies in the collection of similar data also adds to the difficulties of the researchers. For instance, as pointed out above, data relating to Indian industrial sector are available from at least twenty different sources. These alternate sources of data differ in their conceptualization, coverage and measurement procedures. For example, the Annual Survey of Industries (ASI) gives the gross output of industry in value terms where as the the Index of Industrial Production (IIP) gives the index in terms of quantities. Similarly, the ASI covers all companies registered under the Factories Act, where as the IIP covers units registered with institutions, such as the Director General of Technical Development, Coal Commissioner and Textile Commissioner. Consequently one may arrive at different answers to the same question. For instance, to the question how much did Indian industry grow in a particular year, the estimates vary depending upon the source used for estimation. In one such exercise, the divergence among the estimates of growth rates using different sources of data for the year 1978-79 was found to vary between 7.2 percent and 14.1 percent, at the extremes.

Non Accessibility of Unpublished Data

It may be noted further that researchers usually do not have easy access to the unpublished data. Procedural complexities and red tape more often discourage them from making use of such data. Worst yet, potential users do not even have any means of knowing about the available data. If, through personal contacts or by sheer coincidence, they come to know about some unpublished data, it is extremely difficult for them to get access to such data. The dictum of confidentiality and privacy in respect of government held data is so rigidly adhered to by various government agencies, that even interdepartmental exchange of data within the government is difficult. The problem becomes all the more serious when the agency collects data under the provisions of the Collection of Statistics Act of 1953. In denying access to unpublished data they invariably invoke the sanctity of rule 7(1) of the Act which prohibits giving information on individual units. The said rule reads:

"No information, no individual return and no part of an individual return with respect to any particular industrial or commercial concern given for the purpose of this Act shall, without previous consent in writing of the owner for the time being of the industrial or commercial concern in relation to which the information or return was given or made or his authorized agent, be published in such a manner as would enable any particulars to be identified as referring to a particular concern."

It can be seen that the Act does not explicitly restrict the release of raw data to researchers. Government agencies, in fact, use the Act in self defense and to conceal inefficiency in handling data. It may also be appropriate to mention here that while the Act can be invoked by the competent authority to punish or impose a fine if any person "willfully refuses or without lawful excuse neglects to furnish such information... as may be required under the Act", the Act is silent about the obligation of the 'statistics authority' in respect of utilization of the collected information. Consequently, large volumes of collected information continue to gather dust while genuine researchers are denied access to the required data on some pretext or other.

Prospects

A large majority of researchers are working in the universities and research institutions in various disciplines. The universities and deemed universities (institutions which have been given the status of a university), numbering over 150, are financially supported by the government via the University Grants Commission (UGC). The research institutions are financed or are in some way supported by national level organizations in various fields like agriculture, medicine, social sciences, education, science and technology and culture, etc. These national level organizations are, to name just a few: the Indian Council of Agricultural Research, Indian Council of Medical Research, Indian Council of Social Science Research, Indian Council for Scientific and Industrial Research, National Council of Educational Research and Training, etc. All such national level organizations as well as the UGC are financed by the government. As already pointed out, researchers based in these organizations, or institutions supported by them, have quite often little or no access to unpublished government generated data. The published data, on the other hand, serve only a very limited purpose for the scholar. Consequently, researchers have no option but to generate their own data. For want of adequate funds, most of the scholars both in the universities and research institutions end up taking only micro level studies. It is ironic that at least some of the scholars in the universities and research institutions, keen, and qualified to handle large data for macro analysis, must content themselves with only micro studies for want of financial resources. On the other hand vast body of data generated by the government remain largely unutilized. Thus, research activity is seriously hampered not for want of resources, initiative, willingness, skills or expertise in the country, but for lack of perspective and policy for proper storage and sharing of data generated by different wings of the government.

Thus what is needed is not additional output of financial resources for the collection of new data, but a perspective and policy for utilizing data collected by various government organizations. It is necessary to improve communication channels between the producers and users of data so that interested scholars can know what type of data is available, where, and in what form. Various national level organizations may help to bridge this communication gap. Besides, these organizations may try to persuade government agencies to release raw data to genuine researchers. Computerization of govenment data may also greatly facilitate utilization of the existing data. In the absence of a computerized data base, it is extremely difficult for researchers to utilize available data even when these are released by the government. The volume of data is so large, and the structure so complex, that individual scholars with limited funds at their disposal, would find it difficult to put them in proper shape for meaningful analysis.

It should be mentioned that the Indian Council of Social Science Research is making efforts to persuade the CSC, NSSO and other major data producing agencies to, at least, release 'non-sensitive' and unclassified data to interested scholars. The ICSSR is also preparing an inventory of data available in various government agencies and research institutions which will, hopefully, go a long way towards disseminating information to scholars about available data. The ICSSR has also recently financed a project in collabouration with the Bureau of Industrial Costs and Prices, the Planning Commission and the Industrial Development Bank, to develop an Information System on Industrial Data. The System would be developed by pooling industrial data collected by various government agencies and standardizing them as much as possible.

Currently, the government is following a more liberal policy in the use of computers in various government activities. A number of government agencies have already set up computers and transformed their databases in machine-readable form. The National Informatics centre, set up by the Department of Electronics, is also helping various government agencies to computerize their databases. With the help of the National Informatics centre, separate information systems are being developed, containing data on education and manpower, energy, health and demography, industrial resources, agriculture, transport and industry. The ICSSR is part of the NIC network and, therefore, will be in a position to help scholars get access to these information systems, when developed.

Thus the potential of utilizing government generated data in India is enormous, provided proper perspective is developed and policy formulated. Hopefully, given recent developments in government, particularly, in the computerization of government data, it may be possible for researchers to access primary data. Proper utilization of existing data may in turn help government agencies to cut the cost of collecton of unnecessary and irrelevant data. This will, however, require greater competence on the part of the researchers in handling complex data. Recent developments do seem to hold promise but time alone will bear testimony to this.■

Building Regional Data Files in the Federal Republic of Germany

by Edwin Ferger [1]
Zentralarchiv für empirische Sozialforschung
Universität zu Köln

The Zentralarchiv für empirische Sozialforschung, Universität zu Köln, has engaged itself in the creation of the German part of the 'Joint European Time-Series Data Base'. The development of this data base for comparative social science research is a project by the IFDO (International Federation of Data Organizations for the Social Sciences).

In the following paper, I will describe some of our experiences building regional data files, I will point out where we stand now and some problems we face.

Experiences in the Past

Although the Zentralarchiv's primary tradition is to file surveys, there are several election studies in its holdings containing aggregate data; for the most part these are variables of basic demography and occupational structure, as well as political variables.

Many talks with social scientists have shed light on the fact that research interests are diverse. However, there is a strong desire for a common dataset on the lowest possible regional unit. Some regional data files exist, but access is sometimes restricted; data in different files are seldom comparable because they are valid for different demarcations or time points. Data comparability also suffers from boundary or definition changes. Some other problems apply to special circumstances in the Federal Republic of Germany.

During the past years many data files were created in Germany. Unfortunately the data are as heterogeneous as the projects and the underlying research interests from which they result. Although the data files contain precious data, they only have a few variables in common. Knowledge of the existence of data files seems not to be widespread, as seen in the duplication of effort that is usual because of the many authorities and researchers that use large amounts of manpower to create machine readable data sets with more or less overlapping variables.

Usually, existing data files are not easily available. Data sets under the jurisdiction of authorities may theoretically be accessible to the public, however, to get the data is another matter. The transmission of data to a third researcher or research institute almost always involves prolong negotiations or severe restrictions.

---

[1] Presented at IASSIST/IFDO International Conference May 1985, Amsterdam

One of our aims is to build a *data pool* that can be enlarged by the addition of data from different sources. There are some studies already in the Zentralarchiv holdings which contain aggregate data. These sources should be usable for the pool, together with the data we transcribe from printed material to a machine readable format.

At the present time efforts are made to complete the IFDO list of variables as far as possible for the Regierungsbezirk level and later on for the Kreis level.

We try to collect data from other researchers' projects and studies not only for these high levels of aggregation but also for smaller units. The intended combination of data for different regional levels makes data available for multilevel and contextual analyses.

It seems to be reasonable to use data from these sources as the nucleus of the data pool and to complete the gaps that remain in the IFDO list of variables. Most of the variables to be added concern settlement structure, education, income, production, health and household conditions.

When the data are merged, they should be subjected to reciprocal control routines. Regional data are hierarchical in structure across the different levels of aggregation and thus can be subjected to control procedures. These can use sum, line or row percentages. It is impossible, however, to execute checks "totally automatically". Security restrictions and unavailability at lower levels also hinder one from coming in contact with the data. Other data, for example relational data, cannot be aggregated.

From our point of view, control routines are necessary in any case; whether data from existing files are merged or transcribed from printed sources. If machine readable data are merged, one should not start with the premise that existing files are faultless.

Considering the diversity of data files in the university sector, especially their regional and temporal coverage, efforts should be strengthened for the better use of the holdings of the statistical offices of the Länder and the Federal Statistical Office. The development of data banks or statistical information systems in most of these institutions during the last decade seems to open new possibilities for such improvement. Experience will show if these data banks can be used to construct regional data files covering the whole territory of the Federal Republic. There have been some disappointing experiences in the past – the fact that the majority of files have been constructed by transcribing data from printed sources, demonstrates this. On the other hand, the willingness of the authorities to cooperate with the scientific community has increased recently. The data banks and information systems have been established with enormous financial and personal expenditures, and users are now made welcome; a large number of users will mean that their efforts have not been in vain.

Where we stand and the problems we face

In our efforts to transcribe data from printed sources to a machine-readable format, we have built four SAS-files, following the 'List of IFDO variables' (LOIV). We have created one file for each of the three census years: 1950, 1961, and 1970, and another file containing annual data for the variables, area, population, births and deaths. These files have been sent to the NSD. (*editor's note: Norwegian Social Science Data Service*).

Accordingly, three questions arise:

1. Why didn't we obtain a data file with the full set of IFDO variables for 1980?

Unfortunately the latest census in the Federal Republic of Germany was conducted in 1970. About 1980 and for several years thereafter a broad public discussion on data protection took place, which prevented a new census. Only recently the federal government decided to postpone the next census until 1990.

The statistical offfices in our country have annual data, of course, but it has merely been updated by different authorities for various reasons.

Moreover, there is no single, definite time point for updating all variables. Another point to be noted is that many data are registered in connection with special events only (with the exception of a census).

Let us take for example a student who has lived for the rest of his life in the same house in the city in which he attended university. According to the registration office, he will have died at the age of 75 still as a student, assuming no census was carried out in the intervening years and he does not move to another house.

Examples like this highlight the problems of reliability of data based on information updated by administrative authorities only on the occurrence of vital events.

In the future, we will try to procure data of this type for 1980, as well as secure information about reliability limitations.

2. Why has the data been stored in separate files?

The answer is succinct and very important: great changes occurred in the numbers, boundaries and names of the regional units as a consequence of territorial reforms in Germany.

The main problem for the development of a regional data file consists in the instability of regional boundaries over time. From 1968 until the end of the seventies, a far reaching territorial reform took place, which affected not only most communes, but caused changes in regional units at higher aggregation levels as well. The territorial reforms were brought about to reorganize the communal self-administration at the local level in order to create an efficient system of common welfare and social security. Walter Christaller's central-place theory is basic to this concept.

For IFDO purposes, the greatest difficulty is the comparison of variables over time. Territorial reform did not only affect communes as a whole, for example the incorporation of an entire commune into another, but many others were divided, and their parts merged with yet other communes.

Figures 1 and 2 give a sense of the amount of change that has taken place in the last decade. Indeed, it's difficult to find any commune with stable boundaries over this period. The statistical offices of the Länder are able to 'reconstruct' the old commune limits by a "backward recalculation" ("Rückwärtsaktualisierung").

When the data on the newly delimited communes are converted into data of communes under the old demarcation, two different systems must be considered if the new commune was an amalgamation of old communes or parts thereof. Under the old system, all data on the commune to be divided in the course of territorial reform, was transmitted to the commune that incorporated the largest portion of the population within its enlarged limits. The new system focuses on distributing area and

Figure 1: ENTWICKLUNG DES GEMEINDEBESTANDES 1957 BIS 1976

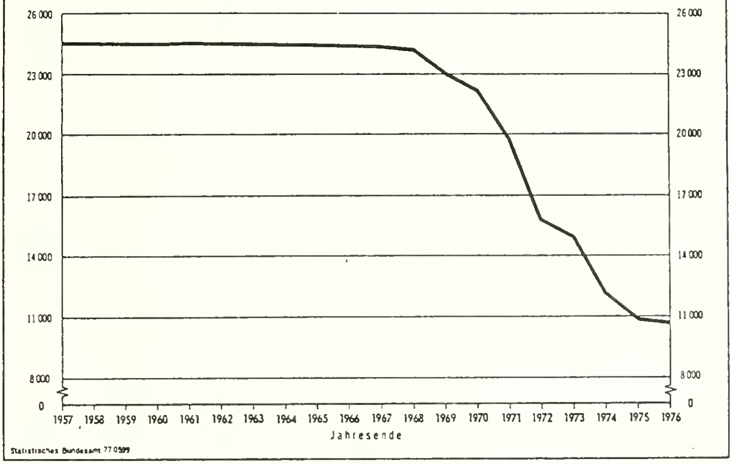
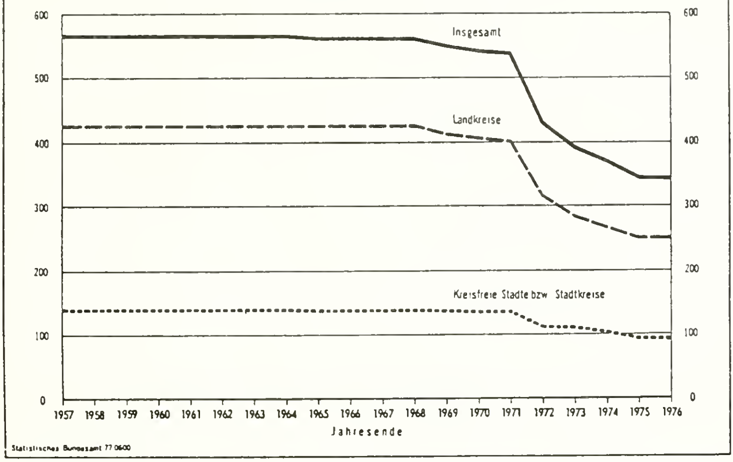


Figure 2: ENTWICKLUNG DER KREISFREIEN STADTE UND LANDKREISE 1957 BIS 1976



Source: Statistisches Bundesamt Wiesbaden (1977): Devölkerungder Gemeinden 1976. Fachserie 1, Bevölkerung und Erwerbstätigkeit, Reihe 1.2.2. Stuttgart: Verlag Kohlhammer. Page 6.

population data exactly. Other data are appropriated to the new enlarged communes according to the percentage of population they received from the old divided commune. A well-known conclusion is as valid here as elsewhere in regional research. The errors, resulting from boundary changes, that occur in comparisons over time tend to vanish at higher aggregation levels.

3. Do we need data which allow comparisons for regional units over time if the regional units have changed?

Changes in the boundaries of regional units are modifications of the units of analysis. If one supports the idea that regions like Gemeinde, Kreis or Regierungsbezirk influence people's lives, one must consider their geographic reality in present as well as earlier times.

Given the many far-reaching changes in regional units by the territorial reform, the aggregation or disaggregation of the old regions within the present-day boundaries means nothing apart from getting a stable scanning pattern. It does not, however, have any influence on people's orientations in their geographic context.

From this point of view, we suggest that the several thousand boundary changes can be tolerated in Germany during the territorial reforms. Regional analysis at different time points are thus rendered possible. Comparisons over time, however, are not only a problem of data availability in Germany but also of adequate research design.■

A Network System For the Flow of African Administrative Information Materials

by Edward Seth Asiedu, [1]
Chief,
Division of Documentation, CAFRAD,
Tangier, Morocco.

Introduction

Nearly three decades ago, the "wind of change" which blew over the African continent left in its wake spates of political agitation, and social movements and pressure groups seeking autonomy from their, then, metropolitan and colonial masters - namely, the British, the French, the Belgians, the Portuguese, and the Spanish.

The granting of independence to these countries brought with it great challenges to their civil services. A consequent change in the variety, number and complexity of government functions has emerged, necessitating a thorough examination not only of the administrative machinery to facilitate the required expansion and modernization for accelerated economic and social development, but also, of training and orientation policies so that civil servants would be better able to meet the challenges posed by the accelerated momentum of change.

Durable administrative reforms and innovations, however, often depend upon research in theories and practices of government. Through administrative surveys, observations, analysis of research questionnaires, interviews, experiments, etc., data concerning inefficiency and corruption can be obtained. Data-collection or fact-finding is the first step towards bringing about administrative reform and creating innovations. If modern government in a developing setting is to accomplish its task and mission effectively and successfully, the administrators who will be required to assume these new governmental responsibilities should have adequate access to organized information to broaden their outlook and training in order to equip them with the capabilities they need.

Studies have shown that administrative sciences are basic to national development. The development of administrative and managerial skills and the application of modern management techniques in an African society faces a lack of adequate and reliable administrative information for making policy and administrative decisions and for analysis of problems. In most cases, the proper administrative information and documentation channels are lacking, thus hindering the transfer of the right amount and kind of information at the right place, to the interested user.

---

[1] Presented at IASSIST/IFDO International Conference May 1985, Amsterdam

Sources of administrative information

What is administrative information? Briefly, let us define the parameters of administrative information. "Administration" may be defined as the achievement of an organizational objective through the use of men, money, materials and machines. "Information" is "a statement that describes an event (or an object or a concept) that helps us distinguish it from others".[2] By extension, therefore, "administrative information" is a meaningful statement about how men, money, materials and machines are managed to achieve the objectives of an organization. In other words, administrative information could be described as information related to administration, designed in general terms, as a set of data necessary to understand institutions, make decisions and to participate in them, in order to produce the goods and services required by the community.

In the African context, however, information on Public Administration is in reality attached to information on "Development Administration". Despite the disagreements on an exact definition of "development administration", most students, scholars, and practitioners involved in the development process in Africa agree that information on the following subjects fall within the mandate of "Development Administration:

namely,

- development planning and plans;
- administrative reform;
- public enterprise management;
- training of administrators;
- research in administrative systems; and
- socio-economic data.

It would also include materials dealing with the administrative and management aspects of the functional areas of the productive sectors e.g. agriculture, industry, eonomic development, etc...

Administrative information sources are usually in the form of pamphlets and report literature, surveys, statistical data, case studies, issuances and other such like documents. A large number of these forms are government publications. Government documents or publications are the living record of the efforts of people to govern themselves. The responsibilities of government extend over many aspects of a citizen's life, work, and leisure and, in the course of administrative control, the agencies of the government collect a vast amount of information. Much of this collected data is of potential value to research workers in economics, social, scientific and technological subjects. Government publications range, in size, from pamphlets to ponderous volumes; and in content, they vary from articles with popular appeal to technical treatises of value. Taken as a whole, they constitute a great library covering almost every field of human knowledge and endeavor. Some of these publications may be transcripts of original records and therefore constitute primary source material in the history of government administration and activities. Others contain accounts by administrative or executive officers on the work and activities under their direction, e.g. annual reports of departments, bureaux, etc... Voluminous series published by other agencies also present statistical pictures of socio-economic conditions and afford bases for measuring social and economic change.

---

[2] Eilon, S. Management control. New York, MacMillan, 1971, pp. 98.

The flow of administrative materials

Data cannot exist if not generated. There is evidence that African governments are becoming more and more prolific in their output of official documents, published and unpublished. The remarkable transformation of postwar Africa, accelerated in the past two decades or so, has created many problems in the field of government publications. There have been extensive constitutional, governmental and administrative changes, ministerial and departmental "re-re-organizations", and alterations in policies governing publishing of official documents. It is becoming increasingly difficult to identify what is available. Whilst in the colonial days, it was easy to acquire the bulk of the official publications from African countries, now it is quite a problem to trace, not to mention acquire, a government document from each other. The reason is obvious - the absence of a plausible, systematic bibliographic control. In-depth research[3] carried out by the speaker in 1977 aimed at studying the generation, production, control, availability and utilization of African Government Documents provides ample information on the situation referred to above.

Another research survey[4], carried out on the provision of library facilities in government departments in a typical African country identified one major draw-back to the improvement of the situation - namely, the lack of financial support and/or importance accorded by African government's to this area. An examination of budget estimates[5] for a four-year period (1971-72/1974-75) revealed some interesting facts. Whilst a few of the Ministries, Departments etc... had huge amounts respectable enough for the establishment, maintenance and running of a library, many others run on votes which look very miserable indeed. In most cases the library votes are hidden under such headings or codes as: "other office expenditures"; "office expenses"; "newspapers";... Worse still, others are found under "toilet papers, stationary and library"; "newspapers and rent of office soap" - a clear indication that libraries, or for that matter the organization of administrative information and literature, are very much frowned upon by the civil service - the very people who need data in their day-to-day occupations. Albeit, it is encouraging to observe that in the last five or so years some improvements have been made in this direction.

The financial situation alluded to above explains the paltry and disorganized condition of the libraries of African government departments. With such an appalling state of affairs, how can proper information be obtained, which information may be vital for the formulation of a major governmental policy? It is pathetic that often most Ministries and Departments lack complete collections of their own publications! Very often, a better organized library outside the governmental system is called upon to satisfy inquiries from Ministries about their own publications! It is amazing to find that a foreign library has everything that African governments have issued - published or unpublished - from the earliest times to the present. African scholars and research fellows have had to travel outside the continent to consult such materials, at great expense to the taxpayer. This situation cannot continue forever; the time is overdue for something more concrete to be initiated.

---

[3] Asiedu, E.S. African governments: the case of West African english speaking countries: Gamiba, Ghana, Liberia, Nigeria & Sierra Leon. Tangier, Morocco, CAFRAD, 1978.

[4] Asiedu, E.S. "Organizing the flow of Administrative Information in a developing setting: a proposal for the Ghana Public Services". In Greenhill Journal of Administration, Vol.1, #3, Oct.-Dec.74, pp.39-56.

[5] See Annexure

The "ANAI" concept

The last twenty five years have witnessed a continued growth of educational and training institutions of public administration and management. Also Civil Service Commission and agencies have been inaugurated to carry out different phases of administrative reform in their respective countries. These are manned by capable African experts who are mainly involved in applying and disseminating their know-how, experience, and knowledge to the African administrative environment. Over the past few years their activities have resulted in the production of vast quantities of documents – often mimeographed reports, government documents and other sorts of publications. These documents describe existing conditions, define goals, describe programmes and projects, and evaluate the consequences of any administrative action and decision. In addition, African administrative science researchers and trainers study administrative processes, generate training materials and publish their findings and conclusions in journals, research reports and other media issued in Africa and elsewhere.

It is quite fair to say that this growing literature of administrative sciences is not yet under adequate bibliographic control and it is exceedingly difficult to know how to obtain access to the information and experiences which have been investigated in print. With this growth of African administrative information comes an increasing demand to know about African experiences in the various aspects of administrative and managerial development.

It is an awareness of this situation, as well as the realization that no one African country can ever hope to become self-sufficient in the face of an expanding universe of administrative information, that the African Training and Research Centre in Administration for Development (CAFRAD) has, in keeping with one of its constitutional mandates, and in concert with African governments, initiated the African Network of Adminstrative Information (ANAI) – a description of which follows:

Objectives of ANAI

ANAI is aimed at bringing together materials from different African perspectives on administrative sciences as they relate to the African continent and bringing together literature which may have been gathered by different African libraries and/or documentation centres, but which provide in their combination a new and much greater resource, and synthesizing and disseminating the collected data at regular intervals and upon request.

Thus, the objectives of ANAI are:

- to improve co-ordination between the existing African administrative information services – i.e. libraries/ documentation centres of institutions and agencies of public administration and management;
- to foster the improvement and building of national administrative information services or systems;
- to supplement local collections by drawing more effectively on African external resources;
- to avoid unnecessary duplication and waste of resources;
- to provide a referral service for enquiries from organizations both inside and outside the African continent;
- to serve more users and improve national,

regional and international access to African administrative and managerial experience and knowledge;

- to make administrative information uniformly available and
- to change the image of the African library from that of a store of books to that of an active and dynamic information centre.

ANAI, as proposed, would imply a degree of diffusion and popularization of administrative information services, a steady increase in the ability to serve at all points of service, and co-operative sharing without constraint of time, distance or form of data. Also, it is intended to minimize the financial pressures which are facing African administrative information activities, and to consider ways of sharing rather than duplicating materials and other resources. It encourages building appropriate local administrative collections to meet immediate needs. It is also devised in such a way as to make readily available distant collections in other African countries. ANAI is aimed also at motivating the utilization of modern miniaturization and computerization techniques in the storage and retrieval of administrative information.

Scope and coverage

ANAI, as seen from the preceeding paragraphs, is a specialized system limited to special categories of users and confined to specific locations. It is anticipated that ANAI will serve a wide variety of users who fall into many categories such as: governmental and other administrative managers; policy makers; educators and trainers; trainees and students; research workers in the domain of the administrative sciences. Also, it is expected that the system will extend its services to any user interested in African administrative and managerial development anywhere in the world, without any restriction on access.

With regard to the subject coverage, ANAI is designed to deal with specialized materials on public administration and management with respect to the African continent. It will encompass data pertaining to topics such as: local government, rural and urban development, administrative reform, project management, financial management, human resources development, etc... As a mission-oriented system, ANAI would have to include relevant multi- and inter-disciplinary information sources in as much as it has a role in the amelioration and improvement of African administrative and managerial development.

The administrative information materials to be fed into ANAI may be recorded in a wide range of media formats: printed forms; audio-visual forms; machine-readable forms (punched cards, magnetic tapes); microphotographic forms (fiches/films). These materials should be mainly generated in Africa, or elsewhere if they are confined to the subject coverage in the African context.

Organizational structure

ANAI is envisioned to function as an African region-wide decentralized information system with a network of (for the time being) three interrelated lines of operation, namely:

1. regional central clearing-house (CAFRAD)
2. national administrative information focal points
3. local libraries and documentaton centres of ministries, departments, bureaux etc...

It is envisioned that a fourth operational line or organizational level could be inserted between regional and national levels to represent a "sub-regional" level, but this development will depend on the magnitude of growth of ANAI in the future.

Brief descriptions of the functions of these levels follow:

- Regional central clearing-house:

  will function as a coordinating agent of the whole network. It will develop and set out technical methods, procedures and standards for the whole network. It will also develop, revise, maintain and unify its indexing language in thesaurus of administrative informatin descriptors; forms for input data, and dissemination channels for output products. It will maintain a data base of information needed, perform miniaturization, microfilming and microfiching of textual input data; and issue guides, directories, abstracts and indexes. It will develop training programmes for the information and documentation personnel of member states and also provide specialist consultancy services to governments as and when required.

- National administrative information focal points:

  These national focal points represent the national participants of ANAI. They would be libraries and/or documentation centres designated by their governments. The role of these national focal points will be to furnish the Regional Clearing-House with local data on administrative and managerial development. These will include input data on institutions and agencies, training programmes, consultants and trainers, training materials, and current research activities which will constitute the data base of the Regional Central Clearing-House.

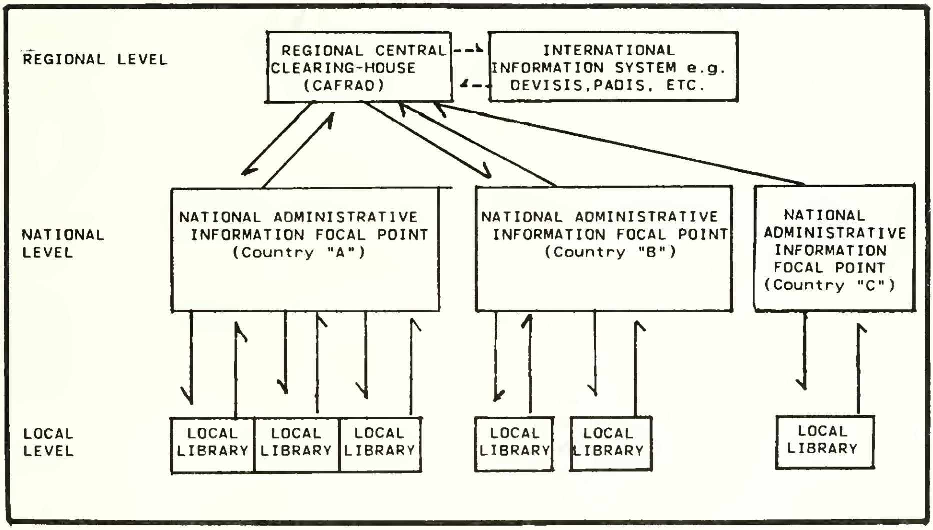
  Each National Focal Point would assist in mobilizing and co-ordinating the national resources and services of administrative information systems. This policy is in conformity with the recent recommendation promoting the creation of national information systems (NATIS) made by the "Inter-Governmental Conference on National Libraries, Documentation and Archives Infrastructure" which was organized by UNESCO, IFLA and FID and held in Paris in September 1974, and also with the UNECA Conference of Ministers' resolution N° 359 (XIV) of the 14th Session held in Rabat, Morocco.

  National Focal Points will be the main suppliers to the Regional Clearing-House of the relevant textual documents, acquisition lists and abstracts. In return, the NCPs will get all ANAI publications, directories, guides, bibliographies, indexes, microfilming services and other dissemination tools pertinent to administrative information.

- Local libraries & documentation centres:

  These will include libraries and other such information units within the civil service organizatons like ministries, bureaux, departments and other state agencies and institutions of public administration and management. These local centres will feed the National Focal Points with their materials and also collabourate with them to identify and supply their input contributions to the Regional Central Clearing-House.

Diagramatically, the organizational structure of ANAI looks like this:


REGIONAL LEVEL
REGIONAL CENTRAL CLEARING-HOUSE (CAFRAD)
INTERNATIONAL INFORMATION SYSTEM e.g. DEVISIS, PADIS, ETC.
NATIONAL LEVEL
NATIONAL ADMINISTRATIVE INFORMATION FOCAL POINT (Country "A")
NATIONAL ADMINISTRATIVE INFORMATION FOCAL POINT (Country "B")
NATIONAL ADMINISTRATIVE INFORMATION FOCAL POINT (Country "C")
LOCAL LEVEL
LOCAL LIBRARY
LOCAL LIBRARY
LOCAL LIBRARY
LOCAL LIBRARY
LOCAL LIBRARY
LOCAL LIBRARY

This kind of organizational relationship could assist in bringing together the African administrative literature under national and regional bibliographic control and this in turn will surely assist and facilitate the identification of resources and access to them. It will also ensure the success of co-ordinating, managing and operating ANAI.

ANAI output

ANAI would ensure output of required data in the required form at the required time to a specific user. This would include the production of specialized guides and directories represented in the different data files stored in the data base, as well as periodic regular abstracts, indexes and specialized bibliographies which represent the current awareness services to users. It will also furnish answers to multiple queries. There will be back-up service of microfiche copies of the full texts of documents identified in the indexes, abstracts and other bibliographies disseminated through the output operation.

Diagramatically this output will look like this:

- Directories & Files
  - Institutions/Agencies
  - Training Programmes
  - Training Materials
  - Consultants & Trainers
  - Current Research Projects
- Output Products
  - Current Awareness Service
    - Monthly Indexes
    - Quarterly Abstracts
    - Specialized Bibliographies
  - Queries – Answers
  - Reprography Services
    - Photostats/Xeros
    - Microfiching
    - Printing

Conclusion

An improved network of African administrative and managerial development needs a high-speed, reliable information system. The conventional means of professional communication, i.e. libraries, provide coverage that is neither complete nor timely, especially as regards research proposed or on-going, as well as decision-making process.

The evolution of a voluntary association of existing libraries and documentation centres of administrative information services in Africa, each preserving its own management autonomy, offers a realistic basis for creating ANAI. Therefore, the ANAI project can be viewed as a regional movement towards increased voluntary cooperation among national partner focal points with varying degrees and modalities of interconnection.

This regional network could be further developed into a world administrative information system that provides current reports on all significant findings, activities and planning affecting any part of the network. For that ultimate goal, the development of ANAI could emerge in conformity and in line with experience gained from the existing international information systems and within the conceptual framework of UNESCO's World Science Information System (UNISIST) programme.

Early this year (January 1985) I.D.R.C. (*editor's note: International Development Research Centre*) of Canada responded to CAFRAD's call for financial support to enable it to implement a pilot project code-named "ANAI EXPERIMENTAL" for a three-year period to test the operational system of ANAI. We are very grateful to the centre for this demonstration of confidence in us. However, there is room for further assistance.

May I, therefore, crave your indulgence to make an appeal, on behalf of CAFRAD: to this august body; to development institutions here present; to other donor agencies world-wide, to collabourate with CAFRAD, to assist CAFRAD, to share with CAFRAD their expertise and financial resources to ensure the germination, growth and fruition of the ANAI seed. □

Library Note Statistics

Department	1971/72	1972/73	1973/74	1974/75	Remarks
Transport & Communication		100.00	6000.00	6000.00	Including Regions
Auditor-General		146.00	3400.00	4400.00	
Trade & Tourism			1700.00	5000.00	
Agriculture	1000.00			3000.00	
Town & Country Planning	100.00		800.00		
National Fire Service	500.00	700.00	760.00	2000.00	
Social Welfare & Com. Dev.		500.00	500.00		
Police Service	600.00	900.00	1500.00	2200.00	Including education instruction materials
Internal Affairs (Central Admin.)					
Prison Service	300.00	400.00	400.00	400.00	Subscription to Ghana Library Board & Prison Library Service
Foreign Affairs		12000.00	8800.00	10000.00	
Posts & Tele-communications					
Local Government (Cent. Admin.)			660.00	4000.00	
NRC (General Administration)		3000.00	3000.00	10000.00	Includes Capital vote of 2730.00
NRC (Press Secretary)			5000.00	5000.00	
NRC (Chieftaincy)		1000.00	3000.00	4500.00	
National Council for Higher Edu.				5000.00	
NRC (Annex)	-	1300.00 .	4000.00	6000.00	
NRC (MSD)	-	6000.00	4000.00	2000.00	
NRC (PIB)	130.00	380.00	1000.00	1400.00	
Regional Organizations	-	1920.00	5220.00	6600.00	No vote for Ashanti & Upper Regions for 1974/75
Parliament House	2000.00	500.00	1000.00	-	
Establishment Secretariat (Training)	800.00	3200.00	1000.00	-	
Public Service Commission	100.00	200.00	200.00		For Newspapers only
Information (Central Admin.)	1250.00	1900.00	1900.00	2000.00	
Information Services Department	-	2800.00	7640.00	6000.00	
Informaion Services (Regions)	-	1400.00	1000.00	3000.00	
Attorney-General's Department	3500.00	32,000.00	8000.00	8000.00	Including Regions
Judicial Services	55,900.00	53,000.00	22,140.00	25,820.00	
Registrar-General's Department	100.00	400.00	1,000.00	1,000.00	
Central Revenue	-	–	1,410.00	10,000.00	
Finance and Economic Planning	70,000.00	60,000.00	? ?	56,500.00	Including 3 other Divisions

Access To Unpublished Social Science Data in Government Departments in the United Kingdom

by D. A. Clarke [1]

Both for the purposes of day-to-day administration and for the formulation of future policy, government departments gather data on a wide range of topics which are potentially valuable for social science research. (Social science is here interpreted very broadly, to include economics and government as well as social studies in the more restricted sense.) A few random examples are agricultural economics and farm profits; consumption of food and drink; public service pay and personnel management; social planning; regional and town planning; population trends; the supply and training of teachers; employment statistics; labour mobility; industrial relations; immigration and race relations; safety, health and welfare; manpower planning; local prisons and penal policy; the police. The records of some other quasi-official bodies, e.g. the research councils and the nationalized industries, must similarly contain data of major value for research.

Though some types of material clearly need to remain closed for a considerable period (e.g. census material and other records of individuals, and data supplied in confidence by business) most of it is not inherently confidential and could be made accessible to *bona fide* researchers long before the expiry of the 30 years after which government records normally become accessible in the Public Record Office (PRO). The problem for the researcher is first to discover what data exists and then to obtain access to it.

In the 1940's and early 1950's the Inter-Departmental Committee on Social and Economic Research (the North Committee) put in hand extensive listings of government material of social science interest (including some unpublished material) held by various departments, which resulted in the publication of a valuable series of six "Guides to Official Sources". The Social Science Research Council, soon after it came into being in 1965, set up a Social Science and Government Committee, which served in a sense as a successor to the North Committee. This latter committee speedily recommended that the North Committee initiative be repeated and the guides brought up to date. The Council therefore awarded a contract to the British Library of Political and Economic Science (at the London School of Economics) to produce a new, single guide to all the unpublished government data likely to be of interest to research workers in the social sciences, and potentially accessible to them, that a small team could locate and list in a limited period. The resulting "Guide to Government Data", published in 1974, reported the material revealed by surveys of 17 departments, carried out over a period of three years from September 1969. It is clearly of the greatest importance that these listings should now be updated, and the remaining departments covered.

Although sheer lack of time necessitated the exclusion of some departments, the absence on other grounds of certain key departments is greatly to be regretted; and it is important that their records should now be surveyed. In particular, the Department of Trade and Industry did not find it possible to cooperate in the enterprise; and, though a great deal of material is published by the department, a guide to unpublished material there remains a most important desideratum for economists. The Foreign and Commonwealth Office was excluded in view of the confidentiality of much of its unpublished material and the practical difficulty of sorting it; those researching contemporary history will nevertheless find it particularly unfortunate that such materials as press releases could not readily be assembled and made available to serious scholars. The records of the former Department of Scientific and Industrial Research contain a vast fund of valuable data relevant to the social sciences in many fields; but preliminary approaches made it clear that a worthwhile survey was impracticable at that time.

---

[1] Presented at IASSIST/IFDO International Conference May 1985, Amsterdam

Though problems of personal privacy and commercial confidentiality will probably always prevent access to recent unpublished data in the Office of Population Surveys and H.M. Customs and Excise, the remaining departments should now be included in a new survey. It seems very possible that in some cases the decision that a department could not cooperate in the earlier project may have been affected by traditional department, or even by personal, attitudes towards public access to official documents *per se*; or by the fact that its library was particularly hard pressed at the time of the enquiry: so that a renewed approach to the department might now be favorably received and a proper survey carries out. It is also to be hoped that the official attitude towards public access is becoming more liberal, and that the policy expressed at the time of the former enquiry by the Civil Service Department, namely that it should "make available as much of the material as it reasonably can within the official limits most generously interpreted" is more widely accepted, so that it will be possible to fund the compilation of a third Guide which would not only update the surveys already published but also reveal important sources of unpublished data held by departments not previously surveyed.

For access to material restricted by the 30-year rule the researcher is normally dependent on the libraries of departments. The amount, type and quality of unpublished material actually made available to a researcher in a library varies considerably from department to department, according to the extent to which it is departmental policy to transfer papers from administrative divisions or branches to its central library. In turn the libraries have varying policies on the admission of researchers, and the extent to which readers, once admitted, are given access to papers, as well as on publication - some departments publish a substantial proportion of the available material.

It may be useful to explain that working papers in a government department are usually held in individual registries attached to the separate divisions or branches which they serve. Although files are reviewed by the department after five years in order to arrange for the destruction of ephemeral material, files containing correspondence and minutes are retained by the department and are not accessible to the public as of right until they are transferred to the PRO after 30 years; and some may be retained by the department after that date on administrative grounds. So if certain information, although not "classified" in the official sense of security classification, is neither readily available in a departmental library open to the public, nor published, nor (as is usual in the case of statistics) provided on payment of a fee to the statistical department concerned, then it generally seems unlikely that under present arrangements members of the public will be able to consult the papers until they have been transferred to the PRO.

It is clear that a substantial proportion of the statistical and other data gathered by government departments is being, or could be, recorded and analysed by automated data processing methods. Both the actual progress made in this area and the scope for further development will no doubt be discussed in the course of this conference. Meanwhile, in 1981 the Committee on Modern Public Records (the Wilson Committee) recommended (para. 423) that "the nucleus of a data archiving centre" should be established as soon as possible; and the Government agreed (Cmnd 8531 para 55) that a feasibility study should be carried out for the establishment of such a centre at the PRO. It is to be hoped that any such centre will work in close cooperation with the Social Science Data Bank of the Economic & Social Research Council.

The advent of data processing techniques should also enable scholars to exploit records more fully and make new connections through the adoption of machine indexing. This should also assist with the hitherto intractable problem of "particular instance papers" (PIPs). These consist of groups of papers, often very numerous indeed, the subject of which is the same in all cases, though each individual paper relates to a different person, body or place. While each individual paper may in isolation be of little importance, when a set of PIPs is taken as a whole and analysed it may form a source from which conclusions as to historical, economic or social trends may be drawn.

In 1979 the appointment of the Wilson Committee provided an opportunity to propose changes which would facilitate the use of government archives by researchers in the social sciences. Some such opportunities which were taken, and others which were missed, are set out below.

In view of the potential importance of data contained in the records of such bodies as the research councils and the nationalized industries, it was recommended that the PRO should make proposals for consistent arrangements for the selection, preservation and public inspection of the records of the research councils and nationalized industries.

Substantial concern had been expressed to the Committee about the "weeding" of records before they are transferred to the PRO. It was felt that coherent, clearly applicable objectives should be established; and that those carrying out this task needed to have an appreciation of the needs of present and future scholars. It was also suggested that some types of "weeded" material might be transferred for preservation outside government departments (e.g. to universities); and that special note should be taken of the importance for research of the working papers of Royal Commissions and Select Committees and the evidence presented to them. The Committee recommended the adoption of a system of "sector panels", including both departmental interests and researchers with experience of the field in question, to advise departments on the official and wider purposes to which departmental records might be put, so that this advice could be taken into account by the "weeders"; but the Government has rejected this recommendation. The Committee made no reference to the suggestion that some "weeded" material might be preserved outside the relevant department: it was perhaps thought that the organization and exploitation of such deposits might require rather substantial expenditures by the recipient organizations.

The Committee made no reference to the suggestion that the press releases issued by Government departments should be computerized; nor to the concern that has been expressed over the rigid application of the "30-year rule" when material on the same topic is already accessible outside the PRO and when material of a routine nature, but of some historical interest, is automatically restricted under the rule.□

New Products

Data User Services Division

Data Developments

BUREAU OF THE CENSUS
U.S. Department of Commerce
Washington, D.C. 20233

Census Offers International Data Base on Diskette

The U.S. Bureau of the Census is now offering its International Data Base (IDB) on diskette for 203 countries of the world. Organized as a series of 94 statistical tables, the IDB contains demographic, economic, and social data for all countries of the world. Each table is fully annotated with information on sources, methods of collection, definitions of terms, methods used in computations, and qualifications of the data.

Each diskette contains tables and notes for a single country. A computer program, written in Basic, is supplied to users to extract individual tables into formats acceptable to software packages such as LOTUS 1-2-3, dDBASE, SUPERCALC, and SPSS[sic]. File documentation describing the contents of the tables and notes is also included. The cost for the first country purchased on diskette, including file documentation, is $15. The price for each additional country is $7.50.

The statistical tables from the IDB are also available in printed copy. When ordering printed copies of statistical tables, the first 10 tables are free and additional tables are $0.25 each.

On-line access to the IDB is available to users within the federal government. Currently, there are 56 organizations within the federal government that have on-line access to the IDB through computer terminals or microcomputers with dial-up capability.

The data contained in the IDB are collected from many sources, including national statistical offices throughout the world, international organizations such as the United Nations, research centres and universities around the world, and federal agencies such as the U.S. Agency for International Development. Data that originate in host countries usually come from census, vital registration systems, surveys, and administrative records.

To supplement these "raw" data, the Bureau of the Census makes estimates and projections on such topics as population, fertility, migration, and vital rates. All these data, together with other reported statistics, are stored in the IDB. Currently, there are over 200,000 records in the data base with updates occurring on a daily basis.

The types of data in the IDB are categorized as: population, urban–rural, age–sex; vital rates, infant mortality, life tables; health and nutrition; fertility and child survivorship; migration; foreign–born and refugee data; provinces and cities; marital status and marriage; family planning; ethnic, religious, and language groups; literacy and education; labour force, employment, income, and gross national product; and household size and housing indicators.

The geographic coverage of the IDB includes 203 countries of the world, usually by urban and rural residence. Selected tables also contain data on provinces and cities. The time coverage is from 1950 to 2025 for estimates and projections.

The IDB is housed at the Department of Commerce Computer centre in Springfield, Virginia, on a Magnuson M80/42 computer. Model 204, a data base management system, is used to store and retrieve data both for printed copies of the statistical tables and for downloading to a microcomputer.

The IDB is maintained by the centre for International Research (CIR), a division within the Bureau of the Census. The centre, comprised of approximately 85 staff members, is engaged in a variety of activities relating to the compilation, analysis, evaluation, and dissemination of international statistics. Staff members who specialize in various countries and subject matter areas are available to assist users regarding many facets of international data. The CIR has produced numerous publications on such topics as summary demographic data for all countries of the world, detailed demographic and socioeconomic data and projections for selected countries, subnational population projections for the U.S.S.R., analysis of Soviet foreign trade, employment in China, and the situation of women in developing regions.

For more information on the International Data Base or activities of the centre for International Research, contact Greg Nowakowski, centre for International Research, Scuderi Building, Room 409, Bureau of the Census, Washington, D.C. 20233; (301) 763-4286.

1980 Statistics on the Immigrant Population Never Before Released!

1980 Statistics on the Immigrant Population

- **Who are the new immigrants?**
- **When and from where did they come?**
- **Where did they settle?**
- **Are they employed? How? Educated? How many are poor?**

A new 1980 census tabulation helps you answer these and many other questions about U.S. foreign-born population. For the first time, users will have access to detailed statistics on immigrants.

The Foreign-Born Population in the United States Microfiche is a "must" for anyone studying the shifting composition of the nation's population. The reports are issued on separate microfiche cards for individual countries, regions and continents. Microfiche for one area is $5.

The following information tells you more about this tabulation and gives ordering information. If you have any questions, contact Customer Services at the Census Bureau (301) 763-4100. Census of Population, 1980: Foreign-Born Population in the United States Microfiche [microform] / prepared by U.S. Bureau of the Census. --Washington: The Bureau [producer and distributor], 1985.

Type of file:

Summary statistics tabulated from sample data.

Universe description:

The foreign-born population in the United States.

Subject-matter description:

These tabulations present demographic, social, and economic data on the foreign-born from the 1980 Census of Population by country of birth and year of immigration. The following is a list of abbreviated table titles:

Table 01. - Age by Sex
Table 02. - Citizenship by Age and Sex

Table 03. – Marital Status by Age and Sex
Table 04. – labour Force Status by Age and Sex
Table 05. – labour Force Status in 1979 by Age and Sex
Table 06. – Industry of Employed Persons by Age and Sex
Table 07. – Occupation of Employed Persons by Age and Sex
Table 08. – Class of Worker of Employed Persons by Age and Sex
Table 09. – labour Force Status of Persons Attending School by Age and Sex
Table 10. – Years of School Completed by Age and Sex
Table 11. – Income and Poverty Status in 1979
Table 12. – Language Usage and Ability to Speak English by Age and Sex
Table 13. – Children Ever Born by Age and Marital Status of Women
Table 14. – Birthplace of Children Under 20 Years
Table 15. – Country of Birth by State and Metropolitan–Nonmetropolitan Area of Residence

Data are provided only for the total United States except for Table 15 which shows foreign–born totals for states by residence inside or outside metropolitan areas. All 15 data tables for a given country, region, subregion, continent or the entire world are on one microfiche (see attached list).

Geographic coverage:

The geographic areas shown in this tabulation generally conform to United Nations (U.N.) classifications because the data were produced partly in response to recommendations by the U.N. and the International Union for the Scientific Study of Population. The U.N. classification does not necessarily conform to the international boundaries formally recognized by the United States government or to the classification shown in the 1980 census reports.

Technical description:

The documentation is in a 24x reduction source document format. The fiche consists of an eye–readable title and maximum of 98 frames. Each frame has a maximum line width of 80 characters and maximum line depth of 64 lines.

The tables are in a 42x reduction COM (computer output microform) format. Each microfiche consists of an eye–readable title, maximum of 206 data frames and two index frames. Each frame has a maximum line depth of 64 lines.

Microfiche availability:

All data tables are contained on one microfiche for a given country or region of the world.
For additional information contact Data User Services Division, Customer Services Branch, Bureau of the Census, Washington, DC 20233. Phone: (301) 763–4100.

The BBC Domesday Project

The BBC Domesday Project: Data Dissemination on videodiscs

by Peter Armstrong [1]

The BBC is undertaking a unique two-year project to produce a national survey of the United Kingdom to mark the 900th anniversary of William the Conqueror's Domesday Book: a national inventory completed in 1086. The material will be published on two interactive videodiscs, containing more than 100,000 photographs and maps, with a gigabyte of digital data used for textual and graphic displays as well as indexes. The first disc is a gazetteer of life in every community in Britain, with text and photographs provided by the local people themselves. Some 15,000 volunteer groups are already collecting this material on their own micro computers, and will be sending it for processing and indexing by the BBC.

The second disc will give the official picture of Britain today using data from Government and academic sources. The interactive video technology will be used to display variables where appropriate against a background map. The aim is to make data, which is theoretically in the public domain, easily available to ordinary people through friendly retrieval software which will then present it vividly on the screen and offer the user the chance to interact with it.

A typical setup could be a local library, where users sit at a standard micro keyboard, connecting a specially developed videodisc player with a color monitor.

Completion of the Project and publication of the Domesday discs are planned for the end of 1986.

---

[1] Editor, Domesday Project

Comparative Projects

Comparative project on class structure and class consciousness

Reports from national projects. Tampere, Finland – April, 1985

Poland

The Polish project will do a pilot in November, 1985 with a sample of 400. Next year they hope to do a survey with a sample of 3,000 individuals. The project needs books and money for travel and to purchase materials. It was suggested that each project contribute $50 to $100 to a fund to be administered by the U.S. project for travel and material needs for Poland.

New Zealand

The project in New Zealand has recently completed a survey of 1000 households. An initial demographic report is available. One of the findings reported that 35% of the population in New Zealand is in the working class. The random sample is unweighted in the demographic report. Merging this data with the comparative data set is projected for the latter part of 1985. They are presently cleaning the data.

Finland

The survey, coding and analysis for the Finnish project was completed in June of 1983. The final monograph, an English translation, was distributed to each project with a request for reactions and comments. The project has received a new grant for two years to work on labour market segmentation. The Finnish data is scheduled to be merged and disseminated with that of the U.S., Sweden, Norway and Canada this summer. They are presently working on recoding some of the variables in preparation for merging.

U.S.

The U.S. data tape and a complete user's guide/codebook is scheduled to be sent to the Inter-University Consortium for Political and Social Research in April, 1985 for dissemination.

The project has outlined at least 5 areas for future analysis:

1. class homogeneity/heterogeneity – families
2. class biography
    a. roles in participation in unions
    b. mobility
    c. friends
    d. unemployment

   e. spouse/family

3. social mobility – U.S., Swedish and Norwegian comparisons

4. the relationship between state employment and class

5. the fall and rise of the petit-bourgeoisie

Australia

The Australian project is not yet totally funded. Queensland and Griffith Universities in a joint effort made the first application for funding in 1983. In 1984 they conducted a pilot of 305 households in Brisbane. So far the project has received $24,000 out of a requested amount of $134,000. A survey of 1750 households is projected to be completed and ready for merging by the end of 1986.

West Germany

Most of the $200,000 the West German project has received from the German Science Foundation has been targeted for contracting out the survey. The pilot/pretesting is to be conducted at the end of April. The survey of a random sample of 2,000 individuals is planned for June, 1985. The data should be coded and returned to the project for analysis by the end of 1985 and projected to be ready for merging the latter part of 1986. Basically, the questionnaire will be a duplicate of the U.S. questionnaire. The sexual division of labour items were initially cut out, but they will try to get them added back in. The anticipated problem with foreigners may be resolved by using bilingual interviewers.

Denmark

The Danish project's original funding application of $70,000 was cut to $50,000 for two years. The survey of a sample of 2,400 individuals is planned for the latter part of 1986. The project has been promised an 80% response rate by the agency conducting the survey. The questionnaire will be similar to that of the U.S. It will take 35 minutes to administer, with all respondents to be above 15 years of age.

Japan

The Japanese project is having problems getting funding. They hope to get funding and have a survey prepared by the end of 1986. A small-scale pilot was recently fielded, which incorporated some class and gender type questions. The project will report the findings in May, 1985.

Canada

The Canadian data is scheduled to be merged and disseminated with that of the U.S., Sweden, Norway and Finland this summer. They are presently working on recording some of the variables in preparation for merging. The project has funding for two more years. They have just begun to analyze the data. Two papers are planned for completion next year.

Norway

The Norwegian data is scheduled to be merged and disseminated with that of the U.S., Sweden, Canada and Finland this summer. The project has funding until the end of 1985. They just finished reports on class, gender, class consciousness and mobility. A class identification paper is to be finished and translated into English this summer. The project has money for two more years to participate in the Scandinavian Comparative Studies Project.

Sweden

Funding for the Swedish project ended two years ago. They have finished their final report. The data is scheduled to be merged and disseminated with that of the U.S., Norway, Canada and Finland this summer. Hakon Leiulfsrud is currently participating in the Scandinavian Comparative Studies Project.

Britain

The British data tape is in process of being cleaned. The project has a well-documented codebook. They hope to begin data analysis as soon as the cleaning problems have been resolved. Four books are planned as products of the data analysis on the following topics:

1. class structure
2. class consciousness
3. attitudes
4. theories of class analysis

Comparative project on class structure and class consciousness merging and dissemination time table

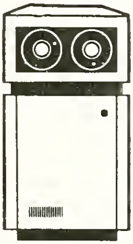
COUNTRY	1985	1986	1987	
U.S.	X			
Sweden	X			
Norway	X			
Canada	X			
Finland	X			
New Zealand		X		
Britain		X		
West Germany		X		
Denmark			X	
Australia			X	
Japan				X
Poland				X

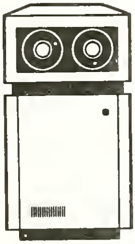
World Election Surveys

Compendia of data collections of certain types on different countries are always of use to those providing service to data analysts. A recent issue of Electoral Studies (1985: 4.2, 159-174) contains a good example of such a compendium, one of academic survey studies of elections around the world. Entries are, in broad terms, routinely repeated academic surveys, national in scope, of regular elections from 1964 to the present. Most of the entered studies are generally accessible through data archives, many of them are members of the International Federation of Data Organizations (IFDO). Access can usually be arranged through the IFDO member archive in your country, which will obtain data on the users' behalf.

Entries in the tabular compendium include year of study; principal investigator; design information (e.g. panel, cross-sectional, etc.); number of cases (respondents); size (number of variables, number of cards); language of documentation; storage (location) archive and related publications. The authors (Paul de Guchtauaire, Director, Steinmetzarchives in the Netherlands, Larry LeDuc, University of Windsor, Ontario, Canada, and Richard Miemi, University of Rochester) promise to up-date this compendium frequently and regularly which should increase its utility.

The Iassist Quarterly would be interested in receiving notice of other such cross-national compendia of data collections which could be included in future issues.





IASSIST

MEMBERSHIP AND SUBSCRIPTION FEES

The International Association for Social Science Information Services and Technology (IASSIST) is a professional association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information systems specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hardcopy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:

REGULAR MEMBERSHIP: $20 per calendar year

STUDENT MEMBERSHIP: $10 per calendar year

Institional subscriptions to the QUARTERLY are available, but do not confer voting rights or other membership benefits.

INSTITUTIONAL SUBSCRIPTION: $35 per calendar year (includes one volume of the QUARTERLY)

Ms. Jacqueline McGee
The Rand Corporation
1700 Main Street
Santa Monica, California
U.S.A. 90406
(213) 393-0411

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIA